

REMARKS

The Office Action dated September 21, 2007 has been received and carefully noted. The above amendments to the claims, and the following remarks, are submitted as a full and complete response thereto.

Claims 1, 5 and 9 are amended to more particularly point out and distinctly claim the subject matter of the present invention. No new matter is added. Claims 1-21 are respectfully submitted for consideration.

The Office Action rejected claims 1-3, 5-7, 9-11, and 13-15 under 35 U.S.C. 103(a) over US Patent No. 6,526,395 to Morris (Morris), in view of US Patent No 6,931,351 to Verma et al. (Verma). Applicants submit that the cited references, taken individually or in combination, fail to disclose or suggest all of the features recited in any of the pending claims.

Claim 1, from which claims 2-4, 16, and 19 depend, is directed to a method of speech recognition. Audio signals are received from a speech source. Video signals are received from the speech source. It is determined if the audio signals can be processed. Based on the detection that at least a portion of audio signals can not be processed, the video signals are processed. At least one of the audio signals and the video signals are converted into recognizable information. A task is implemented based on the recognizable information.

Claim 5, from which claims 6-8, 17, and 20 depend, is directed to a speech recognition device. An audio signal receiver is configured to receive audio signals from

a speech source. A video signal receiver is configured to receive video signals from the speech source. A processing unit is configured to detect if the audio signals can be processed and if so, to process the audio signals. The video signals are processed based on the detection that at least a portion of the audio signals cannot be processed. A conversion unit is configured to convert at least one of the audio signals and the video signals to recognizable information. An implementation unit is configured to implement a task based on the recognizable information.

Claim 9, from which claims 10-12, 18, and 21 depend, is directed to a system for speech recognition. A first receiving means is configured for receiving audio signals from a speech source. A second receiving means is configured for receiving video signals from the speech source. A processing means is configured for detecting if the audio signals can be processed and processing the audio signals if the audio signals can be processed. The processing means processes the video signals based on the detection that at least a portion of the audio signals can not be processed. A converting means is configured for converting at least one of the audio signals and the video signals to recognizable information. An implementing means is configured for implementing a task based on the recognizable information.

Claim 13 is directed to a method of speech recognition. Audio signals are received from a speech source. Video signals are received from the speech source. If the audio signals can be converted into a recognizable format, the audio signals are processed. The audio signals are converted into recognizable information. The video

signals are processed when a segment of the audio signals can not be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. The processed video signals are converted into the recognizable information. A task is implemented based on the recognizable information.

Claim 14 is directed to a speech recognition device. An audio signal receiver is configured to receive audio signals from a speech source. A video signal receiver is configured to receive video signals from the speech source. A first processing unit is configured to detect if the audio signals can be converted, and if the audio signals can be converted, the audio signals are processed. A first conversion unit is configured to convert the audio signals to recognizable information. A second processing unit is configured to process the video signals when the audio signals cannot be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. A second conversion unit is configured to convert the processed video signals into the recognizable information. An implementation unit is configured to implement a task based on the recognizable information.

Claim 15 is directed to a system for speech recognition. A first receiving means receives audio signals from a speech source. A second receiving means receives video signals from the speech source. A first processing means detects if the audio signals can be converted, and if the audio signals can be converted, the audio signals are processed.

A first converting means converts the audio signals into recognizable information. A second processing means processes the video signals when a segment of the audio signals can not be converted into the recognizable information. The video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information. A second converting means converts the processed video signals into the recognizable information. An implementing means implements a task based on the recognizable information.

Applicants submit that each of the pending claims recites features that are neither disclosed nor suggested in the cited references.

Morris is directed to an apparatus includes a video input unit and an audio input unit. The apparatus also includes a multi-sensor fusion/recognition unit coupled to the video input unit and the audio input unit, and a processor coupled to the multi-sensor fusion/recognition unit. The multi-sensor fusion/recognition unit decodes a combined video and audio stream containing a set of user inputs. Fig. 2 of Morris illustrates input interpretation of the system. According to Morris the audio input and video input are processed in parallel. See Fig. 2, refs. 206, 208, 210 and 212. In block 210, visual gestures are captured from the speech input unit. See col. 4 lines 25-44. The Office Action admits that Morris failed to disclose the feature of “detecting if the audio signal can be processed,” “processing the audio signals if it is detected that the audio signals can be processed,” and “processing the video signals if it is detected that at least a portion of

the audio signal cannot be processed.” The Office Action relied on Verma to cure these deficiencies.

Verma is directed to decision making in classification problems. Verma describes classifying samples to one of a number of predetermined classes using a number of class models or classifiers to form order statistic for each classifier. Verma describes that audio and video vectors are similarly processed. The weight for the audio is determined and since there are only two classifiers, the weight for video is determined as a compliment of the weight for the audio as the linear summation of all weights is “1”. The threshold is defined for sample confidence values of audio. The class confidence value if the audio is checked against its threshold. If this test is passed, the audio weight is computed as a constant term and a term which is dependent on the overall confidence of the audio channel. If the test is failed, the constant term changes. See col. 4 lines 41-51 of Verma.

Applicants respectfully submit that the cited references fail to disclose or suggest at least the feature of “processing the video signals based on a detection that at least a portion of the audio signal cannot be processed,” as recited in claims 1, 5, and 9. Specifically, Applicants respectfully submit that Verma fails to cure the admitted deficiencies of Morris.

Verma fails to disclose or suggest that the processing of the video vector is based on the inability to process a portion of the audio vector. As discussed above, Verma merely describes that the video vector is processed in a similar way, and the weights of the audio and video signal are complimentary.

Still further, Verma does not disclose or suggest that there is a determination whether at least a portion the audio signal can be processed and if not, processing the video signal. For example, Verma does not disclose that if the confidence level of the audio signal is “0” then the video signal is processed instead of the audio signal. This is further evidenced in Fig. 2 of Verma, which merely illustrates that the assigning of weights to the audio 110 and video signals 120.

Thus, Applicants respectfully submit that Verma fails to cure the admitted deficiencies of Morris. Thus, the cited references fail to disclose or suggest all of the features recited in claims 1, 5, and 9.

Regarding claims 13-15, Applicants respectfully submit that the cited references fail to disclose or suggest at least the feature of “wherein the video signals coincide with the segment of the audio signals that cannot be converted into the recognizable information.” In other words, Verma is silent with regards to processing the video signals that coincide with the portion of the audio signal that can not be converted. As stated above, Verma merely describes assigning weights to the audio and video signals, and no determination is made as to whether a portion of the audio signal can be converted before processing the video signal.

Applicants further submit that because claims 2-3, 6, 7, 10, and 11 depend from claims 1, 5, and 9, these claims are allowable at least for the same reasons as claims 1, 5 and 9, as well as for the additional features recited in these dependent claims.

Based at least on the above, applicants respectfully submit that the cited references fail to disclose or suggest all of the features recited in claims 1-3, 5-7, 9-11 and 13-15. Accordingly, withdrawal of the rejection under 35 U.S.C. 103(a) is respectfully requested.

The Office Action rejected claims 4, 8 and 12 under 35 U.S.C. 103(a) as being obvious over Morris and Verma, in view of US Patent No. 6,219,639 to Bakis et al. (Bakis). The Office Action took the position that Morris and Verma disclosed all of the features recited in these claims except storing audio and video signals to a destination source and a transmitter for sending the audio signals and the video signals to a destination source. The Office Action asserted that it is well-known in the art to operate biometric identification via a client/server network, where biometric data is stored on a server and biometric data is collected locally and compared to stored biometric data on the server. The Office Action relied on Bakis in support of this assertion. Applicants submit that the cited references, taken individually or in combination, fail to disclose or suggest all of the features recited in any of the pending claims. Specifically, Morris and Verma are deficient at least for the reasons discussed above, and Bakis fails to cure these deficiencies.

As previously discussed, Bakis is directed to recognizing an individual based on attributes associated with the individual. Bakis describes pre-storing previously extracted biometric attributes which may be later retrieved for the purpose of comparison with

subsequently extracted biometric attributes to see if a match exists between the two. See col. 8 lines 47-53.

However, Applicants submit that Bakis is merely a cumulative reference when combined with Morris. The combination discloses comparing biometric data with stored biometric data. Thus, Bakis fails to cure the significant deficiencies of Morris and Verma discussed above.

Based at least on the above, Applicants respectfully submit that the cited references fail to disclose or suggest all of the features recited in claims 4, 8 and 12. Accordingly, withdrawal of the rejection under 35 U.S.C. 103(a) is respectfully requested.

The Office Action rejected claims 16-18 under 35 U.S.C. 103(a) as being obvious over Morris and Verma, in further view of US Patent No. 6,219,640 to Basu et al. (Basu). The Office Action took the position that Morris and Verma disclosed all of the features of these claims except comparing a number of errors detected in the audio signal with the threshold and determining that the audio signals can not be processed if the number of detected errors equals or exceeds the threshold. The Office Action asserted that Basu disclosed this feature. Applicants respectfully submit that the cited references, taken individually or in combination, fail to disclose or suggest all of the features recited in any of the above claims. Specifically, Morris and Verma are deficient at least for the reasons discussed above, and Basu fails to cure these deficiencies.

Basu is directed to audio-visual recognition and utterance verification. Basu describes performing verification by adding score thresholding or competing models. A video signal is processed that is associated with a video source and processing a video signal associated with the video signal. The processed audio signal is compared with the processed video signal to determine a level of correlation between the signals. However, Applicants respectfully submit that Basu fails to cure the significant deficiencies of Morris and Verma discussed above regarding claims 1, 5 and 9.

Further, Applicants submit that Basu fails to cure the admitted deficiencies of Morris and Verma. As discussed above, the Office Action relied on Basu to disclose the feature of comparing a number of errors detected in the audio signal with the threshold and determining that the audio signals can not be processed if the number of detected errors equals or exceeds the threshold.

However, Applicants respectfully submit that Basu is silent with regards to comparing a number of errors detected in an audio signal with a threshold and making a determination that the audio signal cannot be processed based on the comparison. Basu merely describes comparing a processed audio signal with the processed video signal to determine a level of correlation between the signals. Thus, Basu fails to cure the admitted deficiencies of Morris and Verma.

Based at least on the above, Applicants respectfully submit that the cited references fail to disclose or suggest all of the features recited in claims 16-18.

Accordingly, withdrawal of the rejection under 35 U.S.C. 103(a) is respectfully requested.

The Office Action rejected claims 19-21 under 35 U.S.C. 103(a) as being obvious over Morris and Verma, in further view of US Patent No. 5,412,738 to Brunelli et al. (Brunelli). The Office Action took the position that Morris and Verma disclosed all of the features of these claims except determining if the images of the user are detected and indicating to the user if the video image is not detected. The Office Action asserted that Brunelli disclosed these features. Applicants respectfully submit that the cited references, taken individually or in combination, fail to disclose or suggest all of the features recited in any of the above claims. Specifically, Applicants submit that Morris and Verma are deficient at least for the reasons discussed above, and Brunelli fails to cure these deficiencies.

Brunelli is directed to a recognition i.e., identification and verification, system. Acoustic and visual features are integrated to identify people and to verify their identities. However, Applicants respectfully submit that Brunelli fails to cure the significant deficiencies of Morris and Verma discussed above regarding claims 1, 5, and 9.

Further, Applicants respectfully submit that Brunelli fails to cure the admitted deficiencies of Morris and Verma. As discussed above, the Office Action relied on Brunelli to disclose the feature of determining if the images of the user are detected and indicating to the user if the video image is not detected. The Office Action further alleged that a person is notified when his/her image are not detected because an acoustic

indicator does not prompt the user to speak the words, so the absence of a prompt is equivalent to an indication that the video image was not detected. Applicants respectfully submit that the Office Action is inappropriately reading features into Brunelli. First, the features recited in claims 19-21 are positive steps thus an “indication” is not given that the video image is not detected. Further, the Office Action has not provided evidence that the absence of a prompt is necessarily a positive indication that the video image is not detected because the absence of a prompt could be based on any number of factors, does not necessarily flow from the non-detection of the video image. Thus, Brunelli fails to cure the admitted deficiencies of Morris and Verma.


Based at least on the above, Applicants respectfully submit that the cited references fail to disclose or suggest all of the features recited in claims 19-21. Accordingly, withdrawal of the rejection under 35 U.S.C. 103(a) is respectfully requested.

Applicants submit that each of claims 1-21 recites features that are neither disclosed nor suggested in any of the cited references. Accordingly, it is respectfully requested that each of claims 1-21 be allowed, and this application passed to issue.

If for any reason the Examiner determines that the application is not now in condition for allowance, it is respectfully requested that the Examiner contact, by telephone, the applicant's undersigned attorney at the indicated telephone number to arrange for an interview to expedite the disposition of this application.

In the event this paper is not being timely filed, the applicant respectfully petitions for an appropriate extension of time. Any fees for such an extension together with any additional fees may be charged to Counsel's Deposit Account 50-2222.

Respectfully submitted,



David E. Brown
Registration No. 51,091

Customer No. 32294
SQUIRE, SANDERS & DEMPSEY LLP
14TH Floor
8000 Towers Crescent Drive
Tysons Corner, Virginia 22182-2700
Telephone: 703-720-7800
Fax: 703-720-7802

DEB/cqc/jkm